



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/732,809	12/11/2003	Johannes Heinecke	324-163	5874

7590 07/22/2008
LOWE HAUPTMAN GILMAN & BERNER, LLP
Suite 300
1700 Diagonal Road
Alexandria, VA 22314

EXAMINER

SHAH, PARAS D

ART UNIT	PAPER NUMBER
----------	--------------

2626

MAIL DATE	DELIVERY MODE
-----------	---------------

07/22/2008

PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary

Application No.

10/732,809

Applicant(s)

HEINECKE, JOHANNES

Examiner

PARAS SHAH

Art Unit

2626

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --
Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 27 July 2007.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-9 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-9 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☐ Information Disclosure Statement(s) (PTO/SF/ICE)
Paper No(s)/Mail Date _____
- 4) ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date _____
- 5) ☐ Notice of Informal Patent Application
- 6) ☐ Other: _____

DETAILED ACTION

1. This communication is in response to applicant's arguments and amendments filed on 07/27/2007 and 03/18/2008. Claims 1-9 are currently pending in the application. The Applicants' amendment and remarks have been carefully considered, and are moot in view of new grounds for rejection. Hence, the allowed subject matter as indicated previously is withdrawn.
2. All previous objections and rejections directed to the Applicant's disclosure and claims not discussed in this Office Action have been withdrawn by the Examiner.

Change of Examiner

3. It should be noted that the Examiner of record for this Application has changed from Joel Stoffregen to Paras Shah.

Response to Arguments

4. Applicant's arguments see page 10 of applicant's remarks, filed 07/27/2007, with respect to claims 1-10 have been fully considered and are moot in view of new grounds for rejection. The allowance of claims 1-10 has been withdrawn, which was stated in the previous action dated 10/12/2007.

Claim Rejections - 35 USC § 101

5. 35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Claim 9 is rejected under 35 U.S.C 101 because the claimed invention is directed to non-statutory subject matter.

Claim 9 is drawn to a "storage medium" *per se* as recited in the preamble and as such is non-statutory subject matter since the Specification does not provide adequate support to whether the storage medium is program code (see Applicant's Specification, page 4, lines 22-25, where software modules are described). See MPEP 2106.01 [R-5]. Data structures not claimed as embodied in computer readable media are descriptive material *per se* and are not statutory because they are not capable of causing functional change in the computer. See e.g., Warmerdam, 33 F.3d at 1361, 31, USPQ2d at 1760 (claim to a data structure *per se* held nonstatutory). Such claimed data structures do not define any structural and functional interrelationships between data and other claimed aspects of the invention, which permit the data structure's functionality to be realized. In contrast, a claimed computer readable storage medium encoded with a data structure defines structural and functional interrelationships between the data structure and the computer software and hardware components which permit the data structure's functionality to be realized, and is thus statutory.

Claim Rejections - 35 USC § 112

6. The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

7. Claim 1 is rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. The limitation of "... and that increases as the probability of said found second character string..." found on page 4, lines 5 of the claim amendments is unclear as to what that is referring to. There are two possibilities as to which "that" can be interpreted, which is the second coefficient or it is being referred to the score. Furthermore, the limitation of "said determined language" is unclear as to which language is the determined language, it is unclear as to if the determined language is the first language or the second language. Applicant is advised to clarify the claim language. For the purposes of compact prosecution the claim limitations were interpreted to mean the second coefficient increasing and as for the latter limitation the determined language was interpreted to mean the first language for which the score is being calculated.
8. Claims 2-7 are rejected as being dependent upon an indefinite base claim.
9. Claim 9 is rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. It is unclear as to what the Applicant is seeking to claim by the limitation stated. Specifically, it is unclear as to whether a storage medium is being claimed or a storage device, where a claim to a storage device. Hence, for the purposes of compact prosecution, the limitation was interpreted to be a storage medium.
10. The following is a quotation of the first paragraph of 35 U.S.C. 112:

The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the

art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

11. Claim 9 is rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention. The limitation in claim 9, specifically "storage device or storage medium storing coded indicia" is not found in the Specification filed on 12/11/2003. The closest pertinent portion of the specification refers to a computer terminal, see Applicant's Specification, page 4, lines 22-25.

Claim Rejections - 35 USC § 103

12. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

13. **Claims 1-4, 6-9** are rejected under 35 U.S.C. 103(a) as being unpatentable over VAN DEN AKKER (Patent No.: US 6,415,250) in view of DE CAMPOS (Patent No.: US 6,272,456).

14. Regarding **claims 1 and 8**, VAN DEN AKKER teaches a device for automatically identifying the language of a digital text ("automatic language identification system", column 6, line 40), comprising:

means for prestoring (see col. 11, lines 3-7, memory 20, and 30 and see col. 6, line 56-61, where the storage and memory devices used in conjunction with the system) first character strings that occur frequently anywhere respectively in words of a plurality of predetermined languages and characterize said predetermined languages ("probability table 304 includes an entry for every selected word portion 303 that occurs in at least one of the language corpuses 309", column 10, lines 18-20);

means for prestoring second character strings that are atypical anywhere respectively in words of said predetermined languages ("probability table 304 includes an entry for every selected word portion 303 that occurs in at least one of the language corpuses 309", column 10, lines 18-20 and see col. 9, lines 1-16, where variety of corpora are used.);

means for analyzing words extracted from said digital text thereby constructing for each extracted word all character strings contained in said extracted word ("word portions extracted from the input text 301", column 10, lines 39-40) and having lengths lying between one character and the number of characters in said extracted word ("more or less characters may be included in the predetermined number of characters", column 9, lines 22-23);

means for comparing character strings contained in extracted words to prestored character strings in order to determine scores associated with said predetermined

languages ("identification engine 306 searches the probability table 304 for each of the morphologically-significant word portions extracted from the input text 301, summing the relative probability values associated with each language for each of the extracted word portions", column 10, lines 37-42);

means for comparing each of all character strings contained in each said extracted word individually to said first and second prestored character strings of a determined language so that whenever a first character string is found in said extracted word a score associated with said determined language is increased by a first coefficient depending on the position of said first character string found in said extracted word (see column 10, lines 37-42, and FIG. 6, the suffixes are used for scoring, meaning the values are dependent on the position of the characters, since characters from the suffix are used) and whenever a second character string is found in said extracted word a respective second coefficient that is associated with said found second character string (see FIG. 6, "probability table 304 is altered to include predetermined negative values for those word portions which do not appear in a language corpus 309", column 13, lines 62-64) (e.g. The reference shows the comparison of an extracted word to multiple language corpus, which is seen in Figures 6 and 7. hence, corresponding probabilities are increases or decreased based on probable occurrences of the string); and

means for comparing said scores for said text associated with said predetermined languages in order to determine the highest of said scores, which identifies the language of said text ("the largest accumulated relative likelihood value,

provided it exceeds zero, identifies the language of the input text 301", column 10, lines 42-44).

However, VAN DEN AKKER does not disclose that whenever a second character string is found in said extracted word in said extracted word, said score is decreased by a respective second coefficient and that increases as the probability of said found character string in said determined language decreases.

In the same field of language identification, DE CAMPOS teaches whenever a second character string is found in said extracted word in said extracted word (see col. 3, lines 60-67, if the character string is found in many languages, therefore a second character string is analyzed), said score is decreased by a respective second coefficient (see col. 3, lines 65-66, score is decreased if found in many languages) and that increases as the probability of said found character string in said determined language decreases (see col. 3, lines 60-67, score is increased for infrequently appearing strings for the specific language is increased, but if it occurs in another languages score decreases and that increases the probability that the determined language is not the language. Although a second coefficient is not used it would have been obvious to one skilled in the art to add two separate coefficients rather than increasing or decreasing for the objective of discriminating between infrequent sequences (i.e. score (language 1) = $\alpha - \beta$) (see DE CAMPOS, col. 4, lines 62-65)) (e.g. Further, the claimed limitation of the coefficient increasing is evident by the decrease for frequently occurring words in other languages, which entails that a decreasing score lead to a lesser determination that the extracted word came from that language).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to use the coefficient modification of DE CAMPOS in the language identification system of VAN DEN AKKER in order to discriminate languages in identifying languages with infrequently appearing sequence (see DE CAMPOS, col. 3, lines 67-col. 4 lines 1-4 and lines 62-65).

15. Regarding **claim 2**, VAN DEN AKKER in view of DE CAMPOS teach all of the limitations as in claim 1 above. VAN DEN AKKER further teaches that a first character string in an extracted word consists of one of the following character strings: a prefix, a pseudo-prefix, a suffix, a pseudo-suffix, an infix, a pseudo-infix ("word portions containing other types of morphemes or portions of morphemes", column 8, lines 66-67, where "affixes [prefixes, suffixes, infixes] are examples of bound morphemes", column 8, lines 9-10).

16. Regarding **claim 3**, VAN DEN AKKER in view of DE CAMPOS teach all of the limitations as in claim 1 above. VAN DEN AKKER further teaches that said first coefficient of a first character string in said extracted word depends on the frequency of said character string in said determined language ("frequency value indicative of the number of times the selected word portion was found within the corresponding language corpus 309", column 9, lines 36-38).

17. Regarding **claim 4**, VAN DEN AKKER in view of DE CAMPOS teaches all of the limitations as in claim 1 above. DE CAMPOS further teaches that said first coefficient of a first character string in said extracted word depends on the length of said character string ("the language ID program module 36 is looking for the longest match to the test letter sequence of letters appearing in the window", column 13, lines 54-56).

18. Regarding **claim 6**, VAN DEN AKKER in view of DE CAMPOS teaches all of the limitations as in claim 1 above. VAN DEN AKKER further teaches comparator means for comparing each of said extracted words from said text with frequent words in said determined language and initially listed in storage means (see col. 11, lines 3-7, memory 20, and 30 and see col. 6, line 56-61, where the storage and memory devices used in conjunction with the system) so that whenever a frequent word is found in said text said score for said determined language is increased only by a coefficient depending on the frequency of said extracted word in said determined language ("identification engine 306 searches the probability table 304 for each of the morphologically-significant word portions extracted from the input text 301, summing the relative probability values associated with each language for each of the extracted word portions", column 10, lines 37-42) (e.g. Depending on whether word portion is found the probability values are summed increasing the score).

Furthermore, DE CAMPOS teaches increasing the score for one of the languages when the longest match is found in a few languages.

19. Regarding **claim 7**, VAN DEN AKKER in view of DE CAMPOS teach all of the limitations as in claim 1 above. VAN DEN AKKER further teaches the storage means. (see col. 11, lines 3-7, memory 20, and 30 and see col. 6, line 56-61, where the storage and memory devices used in conjunction with the system).

DE CAMPOS further teaches comparator means for comparing each of said extracted words from said text with frequent words in said determined language and initially listed in storage means so that whenever a frequent word is found in said text said score for said determined language is increased only by a coefficient depending on the length of said frequent word ("the language ID program module 36 is looking for the longest match to the test letter sequence of letters appearing in the window", column 13, lines 54-56 and col. 18, lines 26-31, based on length of a word the longer matches are increased in terms of score value).

Allowable Subject Matter

20. Claim 5 would be allowable if rewritten to overcome the rejection(s) under 35 U.S.C. 112, 2nd paragraph, set forth in this Office action and to include all of the limitations of the base claim and any intervening claims.

21. The following is a statement of reasons for the indication of allowable subject matter: DE CAMPOS teaches a score for each language based upon a frequency parameter in the n-gram profiles corresponding to the length of the longest match. VAN DEN AKKER teaches a probability value corresponds directly to the frequency FR.

However, none of the prior art references or in combination thereof teach the coefficient of a first character string equal to $PO(FR + LON)$, as recited in claim 5.

Conclusion

22. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Powell (US 6,157,905) is cited to disclose identification of language using a character set. Paulsen, Jr *et al.* (US 6,704,698) is cited to disclose natural language identification based on word counting. Dunning (US 7,251,665) is cited to disclose finding equivalent character strings relevant to a query. Tong *et al.* (US 7,359,851) is cited to disclose identification of a language of a textual passage using n-grams. Veerappan *et al.* (US 2004/0205675) is cited to disclose document language determination.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to PARAS SHAH whose telephone number is (571)270-1650. The examiner can normally be reached on MON.-THURS. 7:00a.m.-4:00p.m. EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Patrick Edouard can be reached on (571)272-7603. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Art Unit: 2626

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/P. S./

Examiner, Art Unit 2626

07/02/2008

/Patrick N. Edouard/

Supervisory Patent Examiner, Art Unit 2626